Parag Jain

Email: parag.jain@ed.ac.uk
Web: parajain.github.io

Education

University of Edinburgh — Sept 2019 - Current
CDT-NLP PhD, Natural Language Processing, Advisor: Prof. Mirella Lapata

Indian Institute of Technology Hyderabad, India — July 2013 - June 2015
M.Tech, Computer Science & Engg, Advisor: Prof. Vineeth N Balasubramanian — CGPA: 8.92/10
Thesis: Metric Learning for Clustering in Streaming Large-Scale Data - link
Collaborated with Dr. TVS Udaya Bhaskar (INCOIS, India) on Ocean Biome Classification
Awarded A+ (highest possible) for research thesis

Indian Institute of Technology Kharagpur, India — July 2012 - June 2013
M.Tech, Multimedia Information Processing — CGPA: 8.1/10
Completed one year coursework

Jabalpur Engineering College, India — July 2008 - June 2012
B.E., Computer Science & Engg — CGPA: 7.5/10

Industry Experience

IBM Research — Feb 2018 - Sept 2019
Research Engineer

IBM Research — Aug 2016 - Feb 2018
Research Software Engineer (FTH)

Amazon — July 2015 -Aug 2016
Software Development Engineer

Teaching Experience

Tutor, University of Edinburgh — Jan 2021 - Apr 2021
Natural Language Understanding, Generation, and Machine Translation

Teaching Assistant, IIT Hyderabad — July 2013 - June 2015
Numerical Linear Algebra,
Database Management Systems,
Advanced Compiler Design

Travel Grants

- Travel Grant by Facebook AI Research to present our work at EurNLP 2019
- Travel Grant by Xerox Research for attending XRCI Open, Bangalore - 2016
- Travel Grant for attending the Fifth User Interaction Workshop at INCOIS, Hyderabad - 2015

Awards

- Nominated by the University of Edinburgh (one of two) for Google PhD Fellowship 2020
- IBM Second Invention Plateau Award - 2019
- IBM First Invention Plateau Award - 2019
- Equity Award for *outstanding achievement* in research and development. - 2018
  (An individual award selected globally every year by the CEO of IBM.)
- IBM Research First Patent Award - 2018
- IBM Research Certificate of Appreciation (for my contribution in Natural Language Generation Research) - 2018
- Finalist Amazon Idea hackathon, Hyderabad 2016

- Ministry of Human Resource Development Scholarship by Government of India, 2013 - 2015
- Ministry of Human Resource Development Scholarship by Government of India, 2012
- Selected for the prestigious All India Young Jaina Award - 2010.

Publications

Journal

- Jain, P., Lapata, M.
  Memory-Based Semantic Parsing - (Transactions of the Association for Computational Linguistics 2021)
- Laha, A.*, Jain, P.*, Mishra, A.*, Sankaranarayanan, K.
  Scalable Micro-planned Generation of Discourse from Structured Data - (Computational Linguistics Journal 2019)
  VERBALIZING REAL-TIME MATCH INSIGHTS AT THE US OPEN. [US Open Article] [Times of India] [Digital Creed]

Conference Proceedings

- Jain, P., Lapata, M.,
  Conversational Semantic Parsing using Dynamic Context Graphs (EMNLP 2023)
- Puduppully, R., Jain, P., Chen, N., Steedman, M.,
  Multi-Document Summarization with Centroid-Based Pretraining (ACL 2023)
- Perez-Beltrachini, L., Jain, P., Monti, E., Lapata, M.,
  Semantic Parsing for Conversational Question Answering over Knowledge Graphs (EACL 2023)
- Mittal, A., Saha, D, Jain, P., Sen, J., Jammi., M.,
  Bootstrapping Chatbot Interfaces to Databases Overnight (CODS-COMAD 2021)
- Jain, P., Mishra, A., Azad, A.P. and Sankaranarayanan, K.
  Unsupervised Controllable Text Formalization (AAAI 2019)
- Surya, S., Mishra, A., Laha, A., Jain, P., Sankaranarayanan, K.
  Unsupervised Neural Text Simplification - (ACL 2019)
- Agrawal, P., Jain, P., Dalmiya, A., Mittal, A.
  Unified Semantic Parsing with Weak Supervision (ACL 2019)
- Jain, P., Laha, A., Sankaranarayanan, K., Nema, P., Khapra, M.M. and Shetty, S.,
  A Mixed Hierarchical Attention based Encoder-Decoder Approach for Standard Table Summarization (NAACL-HLT 2018, Short Paper)
- Nema, P.*, Jain, P.*, Shetty, S., Laha, A., Sankaranarayanan, K. and Khapra, M.M.,
  Generating Descriptions from Structured Data Using a Bifocal Attention Mechanism and Gated Orthogonalization (NAACL-HLT 2018)

Workshop/Extended Abstract/Demo

- Jain, P., Agrawal, P.
  Unified Semantic Parsing with Weak Supervision (EurNLP 2019)
- Jain, P., Agrawal, P., Mishra, A., Sukhwani, M., Laha, A. and Sankaranarayanan, K.,
  Story Generation from Sequence of Independent Short Descriptions (SIGKDD Workshop ML4Creativity, 2017)
- Bharadwaj, S. *et al.* [and others, including Jain, P.]
  Creation and Interaction with Large-scale Domain-Specific Knowledge Bases (VLDB 2017, Demonstrations Track)
- Jain, P., Balasubramanian, V.N.
  Online Active Metric Learning for Clustering (XRCI OPEN 2016)

Tutorial

- Storytelling from Structured Data and Knowledge Graphs : An NLG Perspective
  Presented at the Annual Meeting of the Association for Computational Linguistics (ACL), 2019

Technical report/Preprint

- Jain, P.
  Unsupervised Metric Learning Using Low Dimensional Embedding
- Jain, P.
  Metric Learning Tutorial

---

* equal contribution

Selected Patents

- Mittal, A.,Jain, P., Saha, D., Sen, J.
  Ontology-Based Automatic Bootstrapping of State-Based Dialog Systems - US Patent App. US20190228068A1
- Khare, S., Jain, P., Dechu, S., Mani, S., Tamilselvam, S.G
  Enabling Tactile Senses in Digital Shopping US Patent App. US20190163272A1
- Jain, P., Azad, A.P Mishra, A., Sankaranarayanan, K.
  Unsupervised Tunable Stylized Text Transformation - US Patent App. US20200034432A1
- Jain, P., Laha, A., Mishra, A., Sankaranarayanan, K.
  Generation of Variable Natural Language Descriptions From Structured Data US Patent App. US20200073944A1
- Sankaran, A., Laha, Lohia, P., Jain, P., Shrivastava, D., Agrawal, P.
  Cognitive Assistant for Co-Generating Creative Content - US Patent Application No. US20200134089A1
- Laha, A., Ekambaram, V., Jain, P., Shrivastava, D.
  Displaying dynamic content on multiple devices - US Patent Application No. US10664217B1
- Chemmengath, S., Jain, P., Laha, A., Krishnan, A.
  Script modification - US Patent Application No. US20200125638A1
- Jain, P., Laha, A., Mishra, A., Sankaranarayanan, K.
  Structured Article Generation - US Patent Application No. US11194816B2

Talks

- (Invited Talk) Conversational Semantic Parsing - Online International. Conference. On. Advances in Physical, Mathematical and Computational Sciences, India - 2022
- (Invited Talk) Introduction to Machine Learning - Christ University Bangalore, 2018
- Metric Learning for Clustering in Streaming Large-Scale Data - IBM Bangalore, 2016
- Introduction to PageRank, IIT Hyderabad, 2015

Service

- Founding team member of TEDxIITHyderabad, 2016
- I server regularly as a program Committee for ACL, EMNLP, NAACL. And have served in past for SIGKDD, Cods-Comad.
- Local coordinator for i.school workshop University of Tokyo, at IIT Hyderabad, 2015
- Student placement representative CSE Department IIT Hyderabad, 2015
- Core member of organizing team Aureoleâ€™10 Engineering Techfest, 2010

Mentorship

- Akash Kumar, Software Engineering intern at IBM (now PhD student at UCSD)
  Project: Tabular Data Summarization
- Aman Sinha, Software Engineering intern at IBM (now Research Assistant at IECL Laboratory, University of Lorraine)
  Project: Scheduled Multi-Task Learning for Semantic Parsing

Relevant Coursework

- **PhD, Edinburgh** - Pragmatics, Controversies in the Data Society, Accelerated Natural Language Processing, Machine Learning and Pattern Recognition
- **M.Tech, IITH** - Hadoop, Advanced Algorithms and Data Structures, Mathematics for Computing, Systems Security

Arts Education

- Diploma (VID I) Violin(Indian Classical Music) from **Raja Mansingh Tomar Music & Arts University, India**.
- Green Belt in Karate from Jabalpur Amateur Karate do association affiliated to **Nepal Karate Do Federation**.

Training and Certifications

- IBM Certified Academic Associate, -DB2 9 Database and Application Fundamentals (ID: PR1587378)
- IBM Certified Deployment professional, -Tivoli Directory Server V6.3 (ID: PR1587378)
- Training on Computer Networking, -CMC Academy, Jabalpur (45 days)